Chapter 2

Robot Mapping

Robot mapping is concerned with developing techniques that enable a mobile robot to construct and maintain a model of its environment based on spatial information gathered over time. Typically, the spatial information stems from directly perceiving the environment through external sensors. In addition, internal sensors like odometry provide information about change of location within the environment. There are, however, many more ways of acquiring spatial information, including external representations such as floor plans, sketches, or written descriptions, as well as direct communication with other robots or with humans.

Most approaches to robot mapping are *incremental* in the sense that a new observation is used to adjust the current spatial model, leading to a new model. The observation is then discarded. The adjustment of the model when new information about the environment becomes available can be seen as a two-step process: First, in the *localization step* corresponding features contained in the new information and in the current model are identified (*data association*) and the robot's position within the model is updated based on the found correspondences (*position update*). Second, in the *map merging* step the spatial information in the model is complemented and updated based on the results of the localization step. The information flow of this general incremental mapping cycle is illustrated in Fig. 2.1. The current spatial model serves as input for the data association and map merging steps and is modified by the map merging step.

Depending on the nature and source of the new information, one of the two suboperations of localization, data association or position update, or the map merging step can be missing. For instance, the robot's internal odometry sensors provide information about the robot's position but not about environmental features. Thus, no data association and no map merging take place. Communicated information on the other hand may or may not provide any evidence about the robot's current position. In the first case, we will speak about *position-related communication*—an example could be information in the form of a you-are-here map—and the information flow would be similar to that of standard incremental mapping. An example of the second case,

J.O. Wallgrün, *Hierarchical Voronoi Graphs: Spatial Representation and Reasoning for Mobile Robots*,
DOI 10.1007/978-3-642-10345-2_2, © Springer-Verlag Berlin Heidelberg 2010

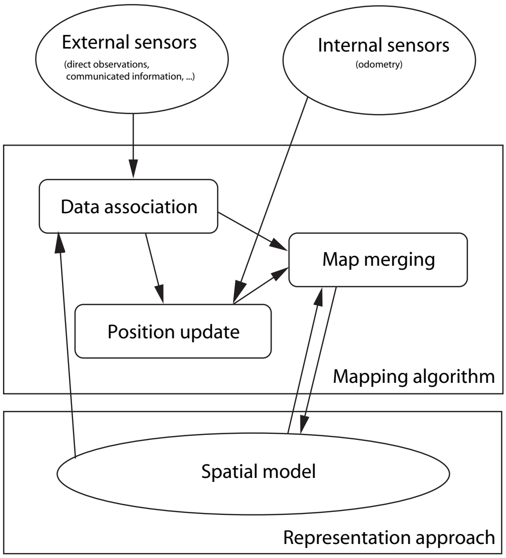

Figure 2.1: Incremental mapping information flow

which we will call *position-unrelated communication*, could be a floor plan which provides a lot of environmental information but no information about the robot's current position. In this case, the data association step would immediately be followed by the map merging step without a position update in between.

Robot mapping is a challenging problem because of the uncertainty inherent in the available spatial information and in the model itself, which always is an approximation of the real world. While it is comparatively easy to localize the robot given an accurate model of the environment (localization with an a priori map) or to generate an accurate model if the exact location is known for every perceived sensor measurement (mapping with known position), the combined problem is very hard because errors in the model and the localization do affect each other. As a consequence, the errors can grow without bounds. The robot mapping problem is therefore also referred to as the *simultaneous localization and mapping (SLAM) problem* (Leonard & Durrant-Whyte, 1991). It is often considered as a state estimation problem and tackled by sophisticated stochastic methods for dealing with the uncertainty, resulting in a research field commonly known as *probabilistic robotics* (Thrun, 2000; Thrun et al., 2005). We provide an overview of the main techniques developed in this field in Appendix A.

Comparing different approaches to the robot mapping problem on a theoretical level is a rather difficult endeavor. The main reason for this is the huge variety of ways in which space can be represented without clear criteria for what makes a good or even optimal representation. As argued in the introduction, over the last several years the robot mapping community has focused on approaching the robot mapping problem from the uncertainty handling perspective. Sophisticated techniques for maintaining probability distributions over high-dimensional state spaces have been developed. However, as a consequence the SLAM problem has mainly been reduced to finding algorithms that compute the most likely model from a sequence of actions and direct sensor observations for an arbitrarily predetermined representation approach.

For the most part this representation approach falls into one of three main classes: occupancy grids, geometric maps, or landmark-based maps. All three kinds of representations describe the environment by providing precise locations of features within a single global coordinate system. This raises the interesting question about why these approaches differ so much from what has been discovered about human cognitive maps (Golledge, 1999). For instance, it is known that human spatial knowledge is fragmented, incomplete, and distorted (Downs & Stea, 1973; Tversky, 1993) as well as hierarchically organized (Briggs, 1973; Hirtle & Jonides, 1985; McNamara, 1986).

In this book, we aim for a more general view on what makes a good or optimal robot mapping approach. Acquiring a spatial model is not regarded as a task per se, but considered in the context of a larger set of competences that are supposed to be realized with the help of the model. Hence, choosing a suitable representation approach becomes part of the problem, and different approaches will be assessed based on how well they support the operations required for different spatial tasks.

In the following, we will proceed by looking more closely at the most important competences that rely on a spatial model of the environment. Based on the set of tasks suggested there, we derive a set of criteria for comparing and evaluating different approaches to the robot mapping problem. Subsequently, we review the robot mapping literature with respect to the spatial representations employed as well as the uncertainty handling methods utilized, and make use of the identified evaluation criteria to compare the different approaches. As a result of this analysis, we will argue that certain promising directions of research have not yet been sufficiently explored. This insight serves as a motivation for the main part of this work, which pursues research in this direction.

In the remainder of this chapter, we will use the term *(spatial) model* or sometimes *map* for the entire spatial information a robot has stored about its environment at a given moment in time. Depending on the considered approach, this can include uncertainty information and different simultaneously considered alternatives. To talk about a single alternative, we will use the term *hypothesis*. *(Spatial) representation (approach)* will refer to the representation formalism used to formulate spatial knowledge in the model. We will further make the distinction between *basic representation approaches*—elementary homogeneous representations—and *organizational forms*—more complex structured representation formalisms that combine different basic representation approaches. Finally, *mapping approach* is used in general to refer to the combination of a spatial representation approach, uncertainty handling techniques, and algorithms employed in a particular implementation.

2.1 A Spatial Model for What?

The problem of acquiring and maintaining a long-term spatial model of an environment cannot be studied in isolation but has to be considered in the context of all the operations that are to be supported by the model. Although the concrete set of operations clearly depends on the specific area of application, we believe that it is reasonable to regard three tasks as fundamental for future mobile robot applications: the tasks of (1) navigation, (2) systematic exploration, and (3) communication about space. We will consider these tasks further below. Other forms of acting in space (e.g., manipulating a set of objects) while being based on spatial properties do not require a long-term model of the environment. Instead, they can be explained as being guided by immediate perception or based on a short-term local spatial model of the agent's proximity. Therefore, we will not consider such tasks in our discussion of environmental models here.

2.1.1 Navigation

Montello (2005, p. 258) defines navigation as a "coordinated and goal-directed movement of one's self (one's body) through the environment." The purpose of navigation is to move oneself in order to reach a particular (known) location. Levitt & Lawton (1990) list three fundamental questions that need to be answered in order to achieve successful navigation: "Where am I?", "Where are other places relative to me?", and "How do I get to other places from here?". Thus, it follows that a spatial representation needs to contain information that supports the ability to localize oneself with respect to environmental features, information about the locations of relevant places, and information that allows us to generate plans on how to reach these relevant places.

Research on animal and human navigation has identified a multitude of navigation behaviors and techniques that contribute to achieving overall successful navigation from low-level guidance-based behaviors over place recognition-triggered responses to higher-level topological and metrical navigation abilities (cf. Trullier et al., 1997). In work on human navigation it is common to distinguish between *locomotion* and *wayfinding* (Darken et al., 1999; Montello, 2005). Locomotion comprises sensor-based movement through the immediate surroundings while avoiding obstacles. Wayfinding comprises all the higher-level cognitive processes involved in navigation like localization and planning. While locomotion can be achieved reactively, wayfinding comprises those abilities that rely on spatial information stored in memory.

From the perspective of operations and spatial competences, we can thus distinguish three main subtasks under the term navigation: *localization*, *path planning*, and *locomotion*. Of these, only the first two require a global spatial model of the environment.

2.1.1.1 Localization

To determine one's location within a model of the environment can mean different things depending on the level of abstraction employed in the model. In an accurate geometric model, it would mean to determine the exact position and orientation of the agent within the global coordinate system used, usually referred to as the robot's *pose*. For a more abstract representation, it could mean, for instance, to establish that one is currently on the town hall square facing north.

In general, it gets easier to localize oneself when more information is included in the model. This is because more ambiguities arise when certain kind of information is not included or details are omitted, leading to increased *perceptual aliasing*. On the other hand, a large amount of detailed information may very well lead to unacceptable space consumption or processing times. However, it is also possible that information required for localization is not distributed homogeneously but mainly provided at certain *distinctive places* (Kuipers, 2000), while navigation between these distinct places is not based on location.

In robotics, three degrees of localization capabilities are distinguished: *Local localization* (or *position tracking*) refers to the ability of tracking your position correctly when your starting position is known. The problem of *global localization* means that the robot should be able to localize when starting at an unknown position. Finally, the term *kidnapped robot problem* (Engelson & McDermott, 1992) is used when the robot is supposed to recover from false localization: The robot wrongly assumes it is at a particular place but is able notice its error and subsequently determines its correct position.

2.1.1.2 Path Planning

Path planning is based on information about how places are connected with each other and about the possibilities to move from one place to another. Hence, it depends on the environment as well as the motion abilities of the robot. Possible connections can be either represented directly, in which case path planning becomes the problem of determining an optimal sequence of connections with respect to a given optimality criterion, or derived based on a description of the traversable space (the *free space*) and the agent's motion abilities.

The resulting plan is a description of how to get to the goal location. This can be specified in many different ways and at different levels of abstraction. It can, for instance, be a detailed sequence of motions to perform, a reactive behavior to execute, or an abstract route description containing references to places and environmental features. In all cases, the plan has to be in line with the motion and sensor abilities of the agent in order to be executable.

2.1.2 Systematic Exploration

Exploration is sometimes seen as a part of navigation, namely when an agent is supposed to find a certain goal location in an unfamiliar environment. This tasks requires a systematic exploration approach to be efficient. In robotics, however, exploration is typically not goal-directed in the sense of reaching a particular goal location. Rather, the purpose is to systematically cover all accessible parts of an environment, either to construct a spatial model for later use or to systematically search or cover the entire environment as in rescue or cleaning scenarios.

The most important aspect of exploration is to keep track of covered parts and parts of the environment that still need to be explored in order to attain complete coverage. In addition, it is often desirable to perform the exploration as efficiently as possible in terms of time or travel distance. In this case, information that supports the decision of which part to explore next needs to be provided by the representation as well.

2.1.3 Communication

Communication about space with other agents can take many forms. First, we can distinguish direct and indirect communication. In the first case, information is directly passed between the agents, typically in verbal form, possibly supported by gestures. In contrast, indirect communication makes use of external representations. Examples are sketches, all kinds of maps (such as topographical maps, you-are-here maps, or floor plans), and written descriptions. The external representation is usually either graphical or verbal.

In a second distinction, one could look at the purpose of the communicated information. Here we could distinguish communication with the goal of conveying information about the environment (like a city map or a verbal description of the city center) from communication with the goal of providing navigational information like route instructions supposed to guide you to a particular place.

Finally, as seen in the beginning of this chapter, communicated spatial information can be position-related or unrelated, depending on whether or not it is adapted to the receiving agent's current position. For example, a city map is position-unrelated as the agent first has to localize within the map in order to use it for navigation, while route directions ("take the third street on the right, then ...") are typically position-related based on the current location of the conversational partner.

In order to communicate with agents that have completely different sensor and motion capabilities, a suitable level of representation needs to be established on which spatial information can be exchanged. To facilitate this, an agent's spatial representation needs to bridge from low-level representations tuned for its own technical abilities to more abstract levels of representation that other types of agents can utilize (Wolter & Richter, 2004).

For communication with humans, a semantic level of description that makes use of relevant human spatial concepts needs to be part of the representation, and the representation needs to be able to mediate between this semantic information and the more sensor-related levels. Anchoring semantic information in the spatial representation has been investigated by Chatila & Laumond (1985) and Galindo et al. (2005).

2.2 Correctness, Consistency, and Criteria for Evaluating Spatial Representations

As mentioned previously, it is not an easy endeavor to compare existing mapping approaches in order to decide which one is most suitable under particular conditions. Assuming that at one point the robot has to make a decision for a single hypothesis, the ultimate goal of the involved mapping algorithms is to derive what we could intuitively call the correct model of the environment from the available information (sensor reading, actions performed, communicated information, background knowledge, etc.), the model that best describes the real environment within the approximation bounds predefined by the chosen representation approach.

However, taking into account the uncertainty of spatial information, we cannot expect that a robot mapping system always comes up with the correct solution. The available information may actually suggest a different state of the environment. Therefore, the goal can only be to compute the most likely or plausible model given the available evidence. Nevertheless, to evaluate the quality of a mapping approach, we would still be interested in its ability to determine the correct model given enough time to gather sufficient data.

Depending on the spatial representation approach chosen, it is common to replace correctness as introduced above by the slightly weaker criterion of the resulting model being *consistent* in the sense that certain crucial spatial properties are represented correctly while other less important ones may not be correct. For instance, again using the example of a geometric model, small discrepancies in the obstacle boundaries can be perfectly acceptable without diminishing the model's usability, while a model in which neighboring rooms overlap would be considered as inconsistent. In the same way, a graph-like representation for a street network with coordinates assigned to the nodes would be considered consistent as long as the graph correctly reflects the topology of the network even if the coordinates do not reproduce the positions of the junctions exactly. Usually, the criterion of consistency is only used intuitively in the robot mapping literature, without giving concrete conditions that a consistent model has to satisfy.

With regard to the spatial representations used in different mapping systems, we propose three general criteria to evaluate and compare different approaches: (1) extractability and maintainability, (2) information adequacy, and (3) efficiency and scalability. We are going to discuss these criteria in the following.

2.2.1 Extractability and Maintainability

It is crucial that a spatial model formulated within the chosen spatial representation approach can be constructed and maintained from the information available to the robot. This means we need to be able to formulate algorithms that take the input information and update the model accordingly. This requires managing the uncertainty in the input information.

Generally, the information required by approaches that model the environment in a low-level sensor-near way seems to be much easier to extract and maintain because no sophisticated processing of the sensor data is needed. The effects of imperfect sensors can be explicitly modeled statistically. On the other hand, more abstract representations may have the advantage that the modeled relations can be more reliably derived from the sensor data, e.g., we might be able to reliably tell that a certain object is between two other objects, while it is much harder to determine the exact shape and location of an object.

Extractability and maintainability of representation approaches can also vary significantly for different kinds of environments. For instance, an approach may only be suitable for well-structured indoor environments or rely on artificial unique landmarks. We will call representation approaches that can adequately represent arbitrary environments *universal*.

2.2.2 Information Adequacy

Similarly crucial is that the chosen representation approach provides all the information that is required for the operations that are supposed to work on the model. The information can be represented either explicitly, meaning that it can be retrieved without further computation, or implicitly, for instance, when the distance between two objects is computed via their coordinates (see Palmer, 1978, for a discussion of implicit and explicit representations). The costs for accessing information implicitly stored in the representation naturally can vary significantly.

Another aspect of information adequacy is that the level of detail is sufficient to support the regarded operations. However, demand for low computational costs and low space consumption (see below) warrants a certain pursuit of sparseness: A representation should not contain superfluous information, information that is required neither for any of the operations nor for maintaining the model itself, and required information should only be represented at a level of detail that is really needed.

2.2.3 Efficiency and Scalability

The ways in which a certain kind of spatial information can be represented are limitless. As a mobile robot typically is supposed to work in realtime, the operations should be as efficient as possible. Efficiency significantly depends on the way the information is represented. Typically, there are trade-offs involved in which one way

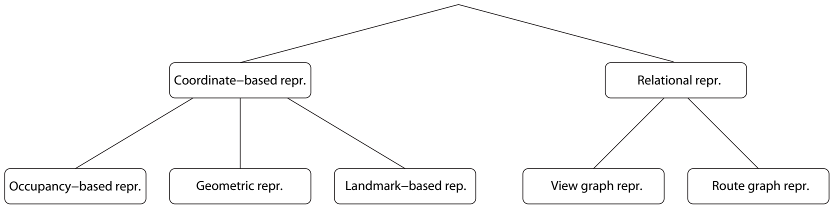

Figure 2.2: Taxonomy of spatial representation formalisms used for mobile robots

of representing things favors a certain operation while making other operations more expensive. A good spatial representation, therefore, would be one which optimizes the overall performance over all operations working on the spatial model, which is rather difficult to assess. However, looking at the complete set of operations, and not only at the efficiency of map construction and map maintenance, will give a much better picture.

In addition to the efficiency aspect, we will use the term *scalability* to discuss how well a representation approach scales with the size of the represented environment. This concerns efficiency of operations as well as space consumption.

2.3 Spatial Representation and Organization

In the following, we review the literature on robot mapping regarding the spatial representation approaches employed. We distinguish between different *basic spatial representation approaches*, which are the elementary representation formalisms to describe an environment in a homogeneous way, and different *organizational forms*, which describe different ways of combining basic spatial representation approaches to form more complex representation structures. Such representations are often referred to as *hybrid representations* (Buschka & Saffiotti, 2004).

2.3.1 Basic Spatial Representation Approaches

Basic spatial representation formalisms mainly differ in two ways: first, in the *basic entities* used to formulate knowledge about the environment, and second, in the way how configurations of the the basic entities are expressed in terms of *spatial relations* holding between the entities.[1] In order to structure this overview on current basic representation formalisms, we use the taxonomy depicted in Fig. 2.2.

At the top level of the taxonomy, approaches are classified based on the way the spatial relations between the basic entities are represented, leading to two main classes: *coordinate-based representations* (often broadly referred to as *metric representations*

[1]As we will see later in this chapter, spatial relations can sometimes be represented rather indirectly in the form of action sequences or motion behaviors.

in the literature) and *relational representations* (comprising, among others, approaches traditionally referred to as *topological maps*). The distinction made here is the following:

Definition 2.1 (Coordinate-based representation)**.** *Coordinate-based representations express spatial relations between basic entities implicitly by providing coordinates for each of the spatial objects within a single absolute coordinate system.*

Definition 2.2 (Relational representation)**.** *Relational representations express spatial relations between basic entities by explicitly stating that a certain relation holds between a certain set of objects.*

The consequences of these two fundamentally different ways to describe configurations will be discussed further below. On the second level of the taxonomy, a distinction is made based on the kind of basic spatial objects used, leading to three subclasses of coordinate-based representations (*occupancy-based representations*, *geometric representations*, and *landmark-based representations*), and two subclasses of relational representations (*view graph representations* and *route graph representations*).

2.3.2 Coordinate-Based Representations

The defining property of coordinate-based representations is that configurations between basic entities are described implicitly through coordinates within a single absolute coordinate system. As a consequence, it is not possible within these approaches to leave the relations between certain entities completely unspecified and thus distinguish between actually perceived relations and derived relations. This leads to the problem that coordinate information needs to be as exact as possible or else global inconsistencies may occur.

On the other hand, if many different spatial relations are required by the operations, coordinate-based representations offer a universal basis from which many different kinds of relations can be derived (e.g., distance, angles, adjacency).

The three most common kinds of basic entities used in coordinate-based representations are cells (used in occupancy-based representations), geometric objects, and landmarks extracted from the sensor information.

2.3.2.1 Occupancy-Based Representations

Occupancy-based representations represent occupied and free parts of space equitably by decomposing space into cells and storing for each cell whether it is (at least partially) occupied or (entirely) free. Typically, the decomposition is independent of the distribution of objects in space and uniform in the sense that all cells have the same shape and size. The dominant decomposition approach employed in robot mapping is the grid map, in which (in the 2D case) a square-shaped raster is used and which allows a simple mapping of locations in the world given in the form of coordinates in a global

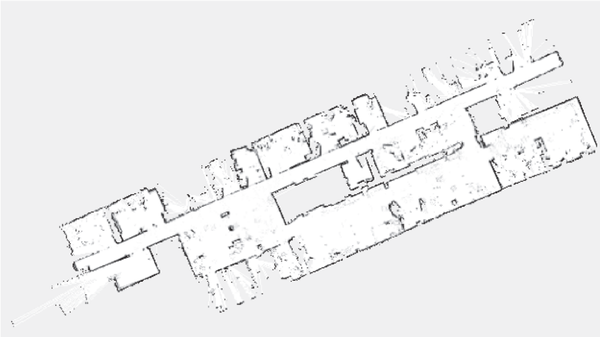

Figure 2.3: An occupancy grid representation. The cells' likelihood of being occupied is represented by their gray values, ranging from white (unoccupied) to black (occupied)

coordinate system to the indices of the corresponding cell in the grid and vice versa. Typically, instead of a binary grid a so-called *occupancy grid* is employed in which the uncertainty about the occupancy state of a cell is represented. In the majority of cases, a likelihood value $l \in [0, 1]$ is maintained for each cell. Figure 2.3 shows an occupancy grid of an indoor environment.

Moravec and Elfes (Elfes, 1989; Moravec & Elfes, 1985) were the first to propose the use of occupancy grid representations for mobile robot navigation and world modeling. They also formulated Bayesian update procedures based on a probabilistic sensor model. Since then, these techniques for learning occupancy grid representations have been refined and combined with advanced probabilistic uncertainty handling methods. Occupancy grids are now used in many state-of-the-art mapping systems (Grisetti et al., 2007a,b; Hähnel et al., 2003a; Thrun, 1998). The idea of occupancy grids has also been adopted to model 3D space (see, for instance, Moravec, 1996).

An extension of occupancy grids called *coverage maps* has been proposed by Stachniss & Burgard (2003a,b) together with suitable probabilistic map update methods. In coverage maps, the degree of coverage is represented as a probability distribution over the interval 0 (completely empty) to 1 (completely occupied) for each cell, resulting in a more accurate description of the environment.

Occupancy grids provide a detailed description of the environment in a sensor-near way, which does not require the extraction of higher entities from the sensor data. That makes it comparatively easy to model the propagation of uncertainty and to develop construction and maintenance methods because no explicit data association is required and incorporating observations (merging step) only involves updating the likelihood values of particular cells. In addition, an occupancy grid representation can slowly adapt to changes in moderately dynamic environments and is universal in the sense that any environment can be adequately modeled.

As an occupancy-based representation preserves most of the spatial information contained in the sensor data, it in principle provides all the information required for navigation and communication. In addition, the high level of detail allows for accurate localization. However, the low-level nature of the represented information without explicit modeling of obstacle boundaries makes many operations rather costly. Path planning can for instance be achieved by value iteration (Howard, 1960; Thrun et al., 1998a), but the search space is rather large compared to other representations. The absence of high-level information or objects complicates the annotation with or the derivation of semantic information as required for high-level reasoning or communication with humans.

With regard to systematic exploration, several techniques have been developed for occupancy-based representations ranging from simply counting the number of times a cell has been scanned, to the identification of so-called *frontiers* between the observed and unobserved areas Yamauchi (1997), to decision-theoretic approaches based on expected information gain (Bourgault et al., 2002).

The main drawback of occupancy-based representations is that they do not scale well to large environments. The high space consumption for larger environments resulting from the fact that the required space depends on the size of the environment and not on the complexity of the environment directly implies strongly increasing computational costs as well. To adequately capture the details in more complex areas a high cell resolution is required which is wasted in less complex areas. Techniques like quad- or octrees (Samet, 1988; Zelinsky, 1992) have been employed to reduce the space consumption problem but can also lead to increased computational costs.

2.3.2.2 Geometric Representations

Geometric representations use parameterized primitive geometric objects, i.e., points, lines, curves, planes, etc. For these objects, we will adopt the term *geom* here. A geometric representation basically consists of a list of geoms describing the boundaries of free space located in a single coordinate system. Most approaches used for mobile robots employ a single kind of geom. Figure 2.4 provides an example of a line-based 2D representation.

Chatila & Laumond (1985) describe an early geometric 2D representation used as part of the world model of their robot HILARE. The representation consists of a set of polygonal objects directly derived from sensor data.

In Crowley (1989), an early approach to construct a model consisting of line segments is described. The line segments are extracted from sonar range data while the robot moves around. Every time a line segment has been detected, it is matched to the model. If a suitable match is found, its parameters are updated accordingly. Otherwise a new line segment is added to the model. The approach is only employed to construct small local models of the robot's immediate surroundings though. In Tardós et al. (2002), techniques for computing a geometric map consisting of point objects (corners) and line objects (walls) are developed.

A lot of work from the area of scan matching has been concerned with computing

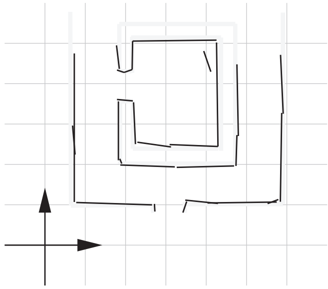

Figure 2.4: Example of a geometric representation: The boundaries of obstacles are represented by line segments

point set 2D representations from laser range data. For instance, Lu & Milios (1997) use a spring-based energy minimization approach to align the individual scans in order to derive a consistent representation. Several groups have developed similar techniques for constructing complete geometric 3D models. Nüchter et al. (2004, 2005) use scan matching of 3D scans together with a global relaxation method to compute a point set representation for complete six degrees of freedom robot motion. A data reduction technique is employed to decrease the number of points from each scan.

An approach for constructing a geometric 3D model consisting of planar surfaces is described in Hähnel et al. (2003b). The approach is based on scan matching to compute a raw 3D model and uses a local search procedure to generate a low-complexity plane model. A different approach described in Liu et al. (2001) focuses on extracting a compact 3D model from a given set of point measurements in 3D space.

Wolter, Latecki, and colleagues (Latecki et al., 2005a,b; Wolter et al., 2004) describe techniques based on shape matching to construct a geometric representation consisting of (generalized) polylines. They employ shape similarity measures to match polylines and complete scans and develop techniques to merge polylines when a new scan is incorporated into the map.

Like occupancy-based representations, geometric representations can represent arbitrary environments as long as geometric primitives are employed that allow us to approximate the shapes of object boundaries sufficiently well (e.g., points, lines, polylines). They are typically much more compact (depending on the geoms used), at least if a merging method is employed to join corresponding entities instead of blindly adding all perceived objects as new objects. However, developing adequate merging schemes is a hard problem and for numerous approaches no merging scheme is described.

As geometric representations describe the boundaries of free space, they are, on the one hand, well-suited for localization and, on the other hand, also allow for path planning. Path planning, however, usually requires the construction of a discrete search space from the representation in order to apply path planning techniques; this produces

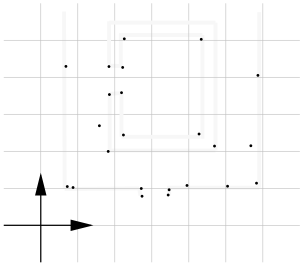

Figure 2.5: Landmark-based representation consisting of point landmarks referenced in a global coordinate system

additional computational costs. Typical approaches here are roadmap approaches, cell decomposition approaches, and potential field approaches (for an overview, see Latombe, 1991).

One disadvantage of geometric representations is that they do not lend themselves very well to systematic exploration: Keeping track of which parts of the environment have been covered requires that the perceived area be explicitly represented and that this description be continuously updated.

With regard to communication, approaches using more complex geometric primitives (e.g., polygons) are usually much better suited than those based simply on points or lines as they allow for describing complete objects which can then be used to attach semantic information. Overall, only certain types of communication can be directly achieved with geometric representations.

2.3.2.3 Landmark-Based Representations

Landmark-based representations represent the world as a set of salient objects (the *landmarks*) extracted from the sensor data. The positions of the landmarks are specified in a global coordinate system (see Fig. 2.5). Both, geometric and landmark-based representations have been subsumed under the term *feature-based representations*, mainly because they can be similarly stored as matrices and lend themselves to the same kind of uncertainty handling techniques.[2] However, from the spatial representation perspective, the distinction makes sense because geometric representations completely model the boundaries of free space, while landmark-based representations focus on specific objects useful for localization and orientation. Besides the positions of the landmarks, additional attributes can be used to discriminate similar landmarks in the data association step.

In the literature, we mainly find simple point-like landmarks that are easy to extract from camera data or range data, or artificial beacons often with a unique ID allowing for unambiguous identification.

[2]The overall approach is also often referred to as *stochastic maps*.

An example for using natural landmarks is the work by Guivant & Nebot (2001). They use tree trunks extracted from laser range data as landmarks to create a 2D map of the Victoria Park in Sydney. Their data set has been used as a benchmark for landmark-based mapping approaches by many other groups (e.g., Montemerlo & Thrun, 2003; Montemerlo et al., 2003; Nieto et al., 2003).

Neira & Tardós (2001) use vision to extract vertical edges corresponding to corners and wall or window frames and project them onto the ground plane to map an indoor environment. In Dissanayake et al. (2001), radar data is employed to map a combination of natural and artificial (radar reflectors) point-like landmarks. The quality of a landmark candidate is assessed by checking its behavior as a stationary point landmark; only stable candidates are incorporated.

Purely artificial landmarks are used, for instance, in Frese (2006b), where indistinguishable landmarks detectable by vision are placed on the floor along the walls. Identification of the landmarks here is based only on the relative positions, taking into account larger constellations of landmarks for global loop closing.

Landmark-based approaches have also been successfully employed in the underwater domain using artificial landmarks in the form of transponders as well as natural features extracted from sonar data (see, for instance, Newman & Leonard, 2003; Williams et al., 2001).

Landmark-based representations are rather compact (depending on the density of landmarks in the environment) and in general scale well to larger environments. They have been extensively used in SLAM approaches, mainly because they lend themselves very well to probabilistic uncertainty handling. Sophisticated techniques for maintaining these kinds of representations in the presence of uncertainty have been developed. Unfortunately, the developed approaches typically lack the universality of occupancy-based or geometric approaches as they are only applicable in environments that provide the required landmarks in sufficient density.

Moreover, landmark-based approaches do not represent the boundaries of free space. This makes them less suitable for path planning or systematic exploration. The only kinds of environments in which this is not problematic are open environments, in which the navigation between the landmarks is not much further restricted by obstacles other than the landmarks themselves. How well landmark-based representations support localization depends to a large degree on the ambiguity of landmarks and their density in the environment. For instance, if the environment is densely covered by landmarks which are easy to distinguish, the localization problem is simplified significantly.

Landmarks play an important role in human wayfinding (Denis, 1997; Lovelace et al., 1999; Sorrows & Hirtle, 1999). Thus, they are well suited to support communication with humans. For instance, they can be used for providing and processing route instructions if it is possible to enable the robot to recognize a set of landmarks similar to those that humans use.

Overall, landmark-based representations are in many cases not adequate as exclusive representations for a mobile robot, but they are well suited to be combined with other approaches.

2.3.3 Relational Representations

Relational representations explicitly enumerate relations that hold between objects. As a result, this allows us to express ignorance because the fact that a relation is not listed means that this relation may or may not hold. In addition, the relational approach allows us to keep track of which relations have been directly perceived and thus are reliable, and what can be deduced from these relations. Moreover, as relational representations usually deal with more abstract spatial relations, ensuring consistency is somewhat simplified. In some approaches the spatial relations are provided indirectly in an action-based form by specifying a sequence of movements or motion behaviors that will move the agent between the locations.

Most relational representations employed in current mobile robots are graph-based representations in which the nodes stand for the basic entities (e.g., views, places, objects) and the edges represent the relevant spatial relations (e.g., adjacency or connectivity). They are often referred to as *topological maps*. We distinguish two kinds of relational graph representations: *view graph representations* and *route graph representations*. In view graph representations the nodes are not directly derived from the environment but are more or less evenly distributed over the free space. Each node is characterized by the view that is available from this particular position. The placement of nodes is based on sufficient perceptual difference with adjacent nodes. In contrast, in route graph representations the nodes are directly induced by the environment. A route graph represents the environment as a network of distinct routes and the nodes stand for distinctive places or particular landmarks encountered along the routes. Often, the route graph representation directly reflects the topology of the free space.

In principle, relational representations may consist of purely propositional statements formulated in a logic-based language. But as graph-based implementations have significant computational advantages because of their analogical nature, this approach is rarely used. However, relational approaches containing graph representations have successfully been embedded into logical frameworks (see for instance Remolina & Kuipers, 2004).

2.3.3.1 View Graph Representations

In view graph representations the nodes are directly associated with the particular sensor input, called a *view*, available at a particular location. A link between two nodes accounts for the fact that both views have been seen in consecutive order, and thus the spatial adjacency of the corresponding locations. The link provides the information the robot requires to move between the two locations. Nodes have to be distributed densely enough so that reliable locomotion between them is possible based on the navigational capabilities of the robot. Construction of the view graph representation

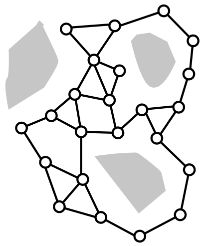

Figure 2.6: View graph representation embedded in the environment

thus means creating a network of nodes that covers the entire environment. Figure 2.6 depicts a view graph embedded within the environment it represents (nodes are located at the positions from which the views have been recorded).

Schölkopf and Mallot introduced the notion of a view graph (Schölkopf & Mallot, 1995) as a kind of minimal model that enables navigation and path planning. They describe a neural network approach for learning the view graph representation from a sequence of views and for using the view graph for navigation in a discrete maze-like world. In later work (Franz et al., 1998), this approach is extended to open environments and implemented on a real robot. The views are given by panorama images from a 360° camera and called *snapshots*. Navigation between views is achieved by a visual homing procedure. Adjacent nodes in the view graph have to be close enough together so that the individual "catchment areas" of the homing procedure overlap. Due to views being 360° images and treated independently of the robot's orientation, and the fact that only sufficiently distinctive views are stored in the graph, it is possible to associate views with particular locations and visualize the view graph as embedded in the environment. In Hübner & Mallot (2007), an extension of the view graph approach involving global position estimates for the nodes is described. Strictly speaking, this approach has to be classified as a hybrid approach combining coordinate-based and relational descriptions.

The ELDEN system described in Yamauchi & Beer (1996) uses a topological map representation that is adaptive, as the edges are annotated with confidence values. Thus, it can to a certain degree deal with changes in the environment such as moved objects, which cause changes in the topology of free space. The places are distributed throughout the environment based on distance from other already established places: If the robot is more than a certain distance away from the last visited place, a new place node is created. Nodes represent regions of space. They are annotated with the global coordinates of their centers and a local occupancy grid representing the geometry of the place. Matching of the grids is used for localization and for hill-climbing to the center of the region. The edges are annotated with the heading information that is updated each time the edge is traversed. The system relies on accurate dead reckoning and has only been evaluated for small environments.

Duckett & Nehmzow (1999a,b) report on experiments with a view graph repre-

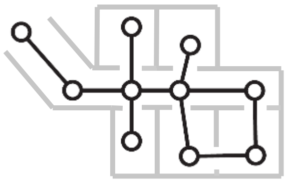

Figure 2.7: Route graph representation of an indoor environment

sentation that is very similar to that of Yamauchi and Beer. The system described in Duckett & Nehmzow (1999a) learns the topological map from sonar, infrared sensors, and compass readings. The nodes are annotated with local range data derived from the sonar measurements. Relative distance and direction information is stored at the edges. A neural network detects open areas and stores them as predicted places for further exploration. For localization, the approach requires global coordinate estimates, which are computed using a spring-based relaxation approach.

View graph approaches are representations very close to the actual sensorimotor experience of the agent but abstract the continuous world into a discrete representation. As such they are rather universally applicable as long as the sensor information is rich enough so that perceptual aliasing does not become a problem. Also, complementing the representation (merging step) is straightforward, and the data association problem is at least reduced.

View graphs provide the information for successful navigation, but the resulting paths tend to be suboptimal, especially when exact homing is required after every step. Their scalability depends on the density of nodes required. The downside of the representation is that almost no structural information about the environment is represented and that the model varies depending on the sensor and motion abilities of the agent and depending on the starting position. As a consequence, the representation is not very well suited for systematic exploration and communication.

2.3.3.2 Route Graph Representations

The route graph concept has been introduced in Werner et al. (2000) as a general model for environmental knowledge gained by integrating route information into survey knowledge by humans and animals, and also by artificial agents. We adopt it here for graph representations in which the nodes stand for distinctive places induced by the environment. The edges reflect distinctive paths connecting these places, allowing travel from one place to another. Figure 2.7, for instance, shows a route graph for an indoor environment in which the nodes correspond to rooms and junctions and the edges correspond to doorways and hallways. Route knowledge is generally assumed to play an important role in the development of human representations of large-scale space (Siegel & White, 1975).

The TOUR model by Kuipers (1978), which is proposed as a psychological model of human common-sense knowledge of large-scale space, describes a representation of a street network environment consisting of places, paths, and regions. It can be learned from abstract route descriptions consisting of sequences of turn and move actions. The place and path descriptions represent the environment as a route graph.

In Kuipers & Byun (1988), these ideas have been transferred to the domain of mobile robots operating in indoor environments. Reactive control procedures for hill-climbing to locally distinctive places (represented by the graph nodes) in the environment or for moving through the environment form the basis for abstracting the environment as a discrete graph structure. The edges stand for control procedures that need to be executed in order to move the robot between the nodes connected by the edge (e.g., following the midline of a corridor). The applicability of particular hill-climbing or control procedures is derived from the properties of the immediate surroundings of the robot. The approach explicitly allows for storing local (geo)metric information for the nodes and edges. An exploration agenda is kept listing nodes with directions that still need to be explored. Localization is based on matching the local description of places and employing a topological *rehearsal procedure* that validates the hypothesis that two places correspond by comparing the results of traveling to neighboring nodes.

These ideas have been further elaborated and refined within Kuipers's framework of the *Spatial Semantic Hierarchy* (SSH) (Kuipers, 2000; Kuipers & Byun, 1991; Kuipers & Levitt, 1988; Remolina & Kuipers, 2004). The SSH contains a topological level of representation that is derived from a sequence of views and actions. Recent implementations of the topological level of the SSH (Kuipers et al., 2004; Modayil et al., 2004) directly employ the Voronoi graphs (see below) of the environment and an extension for open spaces (Beeson et al., 2005) to derive and identify places.

Levitt & Lawton (1990) describe an approach to environmental modeling for open outdoor environments with landmarks. This approach contains a symbolic level of representation which can be interpreted as route graph representation. The landmarks create a natural partition of the environment into regions which can be identified based on the currently visible panorama of landmarks.[3] Adjacent regions share the same panorama with one exception: The orientation of the triangle formed by a random position in the region and the positions of two particular landmarks is reversed. An undirected graph that reflects the adjacency relation is used for qualitative path planning. Navigation between adjacent regions can be achieved by simply crossing the line connecting the two landmarks that make up the difference in the panorama.

Mataric (1992) proposes a distributed map representation that is completely integrated into a system based on the subsumption architecture (Brooks, 1986). The nodes correspond to wall or corridor landmarks observed while exploring the environment using some boundary-following behavior. Nodes experienced consecutively are linked. Localization and path planning are realized by spreading activation within the

[3] An error in this localization approach based on cyclic ordering of landmarks has been pointed out and corrected by Schlieder (1993).

network of nodes, but closing cycles requires at least coarse position estimates for the nodes.

In many approaches, the route network is extracted from the geometry of the environment. One way to achieve this is to follow the idea of retracting free space onto a set of one-dimensional curves called a *roadmap* (Latombe, 1991). One such retraction is provided by the generalized Voronoi diagram (GVD) (Lee & Drysdale III, 1981; Okabe et al., 2000) or the related idea of the medial axis transform (Blum, 1967).

Choset and colleagues (Choset & Nagatani, 2001; Nagatani et al., 1998) directly use the generalized Voronoi graph (GVG), the graph abstraction of the GVD that contains nodes for meet points and edges for the Voronoi curves, as the graph representation of the environment (see Chap. 3 for more details on GVDs and GVGs). The robot navigates and explores the environment using sensor-based control laws that allow for driving along the edges of the GVG. Nagatani & Choset (1999) propose a modified structure called the reduced GVG in which certain kinds of unstable nodes referred to as weak meet points are removed. However, the problem of instabilities is only reduced by this method, and more advanced techniques are needed, as we will discuss in Chap. 4. Choset and colleagues also describe an extension of the GVG into higher dimensional space (>2 dimensions) called the hierarchical GVG, assuring that the crucial properties of the GVG (especially connectedness) are preserved (Choset & Burdick, 2000; Choset et al., 2000).

An alternative to the retraction approach has been described by Chatila & Laumond (Chatila & Laumond, 1985) and Thrun & colleagues (Thrun, 1998; Thrun et al., 1998a). In both approaches a route graph-like representation is derived from a global coordinate-based representation (a geometric one in the former case and a grid-based one in the latter) by first partitioning free space into regions and then capturing the adjacency relation of the regions in the graph structure.

Obviously, route graph representations are tailored to path planning. The structure directly provides a comparatively small search space that can be searched via standard graph searching techniques like Dijkstra's shortest path algorithm (Dijkstra, 1959) or A* (Hart et al., 1968). In contrast, localization is often mentioned as a problem of route graph representations. As theoretically investigated in Dudek et al. (1991) and Rekleitis et al. (1999), topological rehearsal alone is not sufficient; at least one marker is required to guarantee correct localization and map construction. However, in practice providing additional annotations that can be used to distinguish nodes and edges can improve localization within the graph structure significantly.

One advantage of typical route graph representations is that they directly facilitate systematic exploration. To cover the entire environment it is sufficient to trace all edges of the graph, which can be easily achieved by keeping an account of untraversed edges for each node. Furthermore, a representation of the route network of the environment which explicitly represents decision points is very useful for route-based communication (e.g., generating a route description for or interpreting a route description of a human). The compactness of the representation also makes it a good candidate when

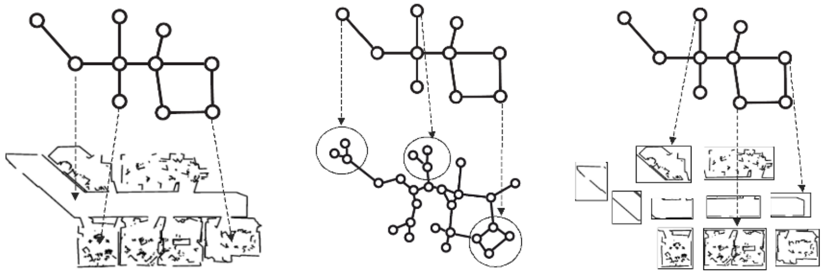

Figure 2.8: Different higher forms of organization: an overlay representation, a hierarchical representation, and a patchworks map

a complete map should be exchanged between multiple agents. In many approaches, places represented by nodes correspond to important concepts such as rooms, which makes it easy to anchor semantic information in the representation.

The main challenge for route graph-based representation approaches is to develop robust methods to construct and maintain the representation. Here, the higher level of abstraction becomes a problem as it is much harder to formulate suitable stochastic models, and the lack of precise localization raises additional problems.

Finally, for route graph representations to be applicable, the environment needs to possess a clear route structure. Open spaces or very unstructured environments do not lend themselves to route graph-based modeling. For structured environments, compactness and the fact that arbitrary additional information can be attached as annotations to the graph structure make route graph representations a promising approach.

2.3.4 Organizational Forms

The previous section provided an overview on basic representation formalisms used for environmental modeling, distinguishing five main classes which show different and often orthogonal strengths and weaknesses, as pointed out by Thrun (1998).

Consequently, people have combined different representation formalisms into more complex forms of organization. We will review these approaches in the following, distinguishing three main forms: *overlays*, *hierarchical representations*, and *patchworks*[4] (see Fig. 2.8). These main forms of combining basic representation approaches can themselves be combined in order to form even more complex forms of representation (cf. Sect. 2.3.4.5).

When reviewing different organizational forms in the following, our main focus is on which strengths and weaknesses are preserved and how the individual representation approaches cooperate. Buschka et al. (Buschka, 2005; Buschka & Saffiotti, 2004) distinguish two modes of cooperation between different representations in a

[4] A classification similar to ours is derived in Buschka (2005).

combined representation which we will adopt here: *Injection* allows us to perform a task within one basic representation based on information from another representation which would otherwise not have been possible to perform (e.g., localization information in a geometric representation is used to localize the robot within a route graph representation). *Synergy*, on the other hand, occurs when two basic representations can be used for the same task but the performance is increased by transferring information between the representations (e.g., localization in the route graph representation alone is possible but the result is improved by incorporating information from localization on the geometrical level).

2.3.4.1 Plain Representation

For completeness, we introduce the term *plain representation* for approaches that only use a single basic representation formalism to describe the environment. Most of the approaches listed above fall into this category, though some of them are actually parts of more complex organizations, as will be pointed out below.

2.3.4.2 Overlays

Overlay representations feature multiple layers of representation. Each layer employs its own representation formalisms and covers the entire environment. The layers are linked in a way that allows us to achieve injection or synergy effects.

Chatila & Laumond (1985) propose an overlay representation consisting of a geometric level and a route graph representation. The route graph representation is actually a hierarchy of graph representations. Its lowest level is directly derived from the geometric representation. Both layers evolve simultaneously while the robot explores the environment.

A similar example is the approach described by Thrun (1998), in which the fundamental representation is a classical occupancy grid used for mapping the environment. Once the environment has been mapped, a topological graph representation is derived from the grid, in which the nodes correspond to regions that form a partition of the free space. Thus, three layers are used in this approach: the occupancy grid representation, a geometric representation describing the regions, and a route graph representation.

Overlay representations typically allow for choosing the best layer of representation to perform a particular operation. This happens at the cost of additional computational efforts to maintain multiple representations and keep not only each layer consistent by itself but also the layers consistent with each other. Typical coordinate-based and graph-based overlays alleviate the problems of the relational representations by improving localization and maintaining the representation either through injection or synergy. However, the disadvantages of the coordinate-based approaches like high space consumption and computational costs typically remain because a global consistent coordinate-based representation still needs to be constructed. Overall, these kinds of overlay approaches tend to not scale well to larger environments.

2.3.4.3 Hierarchical Organization

Hierarchical representations, like overlays, consist of multiple layers, all covering the entire environment. The difference is that here all layers use the same representation formalisms but represent the environment at different levels of granularity. The lowest level provides a rather detailed image of the environment while the higher levels are much coarser or more abstract. In the literature, we can find examples of both, hierarchical coordinate-based representations and hierarchical relational representations.

Fernández & González (1997, 2001) define a general framework for a hierarchical graph-based representation of space called the *AH-graph* (annotated hierarchical graph). The AH-graph consists of multiple graph layers expressing structural information between objects or places. Nodes on higher levels correspond to subgraphs on the level below. Non-structural information like perceptual information about objects or locations or information about the type of structural information described by the edges is stored in form of annotations to the graph structure. Path planning within an AH-graph by hierarchically refining paths and conditions under which this approach yields optimal paths are discussed in Fernández & González (1998).

In Fernández & González (2001) and Fernández-Madrigal & González (2002) this framework is extended to a multi-hierarchical model in which multiple graph-based hierarchies are combined into a single description of the environment. The goal here is to allow for choosing the most appropriate hierarchy for a given task.

In Remolina et al. (1999), the AH-graph model is adopted to describe an extension of the topological level of the SSH, in which places are hierarchically organized into regions. Different approaches for grouping places are discussed, from fully automatic criteria to grouping based on user interaction.

The quad or octree representations mentioned in Sect. 2.3.2.1 can be seen as hierarchical occupancy-based representations of space. However, the tree data structure does not allow for operating exclusively on a chosen level of granularity, as information within one level is only linked via the higher levels. A layered approach as in the graph-based representations listed above would result in very high space requirements, making occupancy-based representations less applicable to hierarchical organization.

Hierarchical organization of information allows us to represent information at different levels of granularity and to either choose the appropriate level for a given task or perform tasks (e.g., path planning) hierarchically by switching to a finer or coarser level as required, increasing the overall efficiency. Especially with regard to communication between heterogeneous agents, this way of bridging from low-level individual levels of representations to more abstract and general levels of representation seems very suitable.

The downside of the hierarchical approach is that additional effort is required to maintain the connections between the layers and keep the individual levels of representation consistent with each other. Also, in most approaches constructing the hierarchy is at least partially based on user interaction. Methods to derive high levels of abstraction autonomously still need to be developed.

2.3.4.4 Patchworks

In patchwork representations, subregions of the environment are represented individually by so-called *local maps* using a single representation formalism, and each using its own frame of reference. The local maps are then related on a higher level to form a global representation. In the literature, the most common forms of patchwork representations are local coordinate-based maps related in a global graph-like relational representation. In addition, we can find approaches that employ local coordinate-based maps related in a global representation which is also coordinate-based.

Local Coordinate-Based and Global Relational Representation. Many authors have proposed representations consisting of local coordinate-based maps linked together on a global topological level. The local maps can be linked either to nodes or edges on the topological level. In some cases, the local maps completely cover the environment, while in others they only describe important areas.

Simhon & Dudek (1998), for instance, employ local geometric maps linked to the nodes of a topological graph representation and discuss how the local maps should be distributed based on a given task (e.g., navigation). The edges linking the nodes on the topological level correspond to control strategies for moving the robot between adjacent local maps.

Yeap & Jefferies (1999) describe a computational model of cognitive mapping in which local line-based geometric maps are constructed, bounded by so-called exits which are identified based on occlusion in the range data. The local maps are disjoint but cover the entire environment. On the topological level, adjacent local maps (those that share at least one exit) are connected by links and links can be enriched with different kinds of information about the connected exits in the local maps. Means for recognizing already visited local maps are discussed but the overall approach relies on a robust detection of the exits independently of the side from which they are approached. This can be hard to achieve in practice, especially in cluttered environments.

In Lisien et al. (2003), a patchwork representation called *Hierarchical Atlas* is proposed which combines local landmark-based maps with a global topological representation. The topological map consists of the reduced generalized Voronoi graph (Nagatani & Choset, 1999) which allows for path planning and systematic exploration. The local landmark representations are attached to the edges of the graph. They support localization within the topological map by disambiguating edges as well as precise localization within the region corresponding to an edge.

Local Coordinate-Based and Global Coordinate-Based Representation. Some mapping approaches describe or can at least be interpreted as using local coordinate-based representations located within a global coordinate system. For instance, scan matching approaches that store scans together with position and orientation estimates of the scans' origins fall into this class (Lu & Milios, 1997; Nüchter et al., 2004, 2005), although often enough the finally intended representation would merge all measurements into a single frame of reference.

Another approach is the work by Chatila & Laumond (1985), in which geometric representations of individual objects, each described within its own reference frame, are related in a global coordinate system. Furthermore, certain kinds of landmark-based mapping approaches employ submapping strategies in order to reduce the computational burden of uncertainty handling (cf. Sect. 2.4.1.2).

A general framework for patchwork approaches is the Atlas framework described in Bosse et al. (2003). The nodes in a topological map stand for local coordinate-based maps (e.g., landmark-based or geometric maps) and edges represent adjacency between the local maps. In addition, the edges represent the coordinate transformation between the local maps. Local maps and global relations between them are both modeled in a stochastic framework. Loop closing is performed based on matching local maps.

Overall, patchwork maps are mainly used to restrict the problem of increasing uncertainty to areas of manageable size, which are then modeled within their own frame of reference, and to reduce computational effort by using a kind of divide-and-conquer approach that avoids simultaneously tracking the relationships between all represented objects. However, the problem of dealing with unbounded uncertainty accumulation still arises when relating the local maps consistently on the global level (e.g., when correctly closing a loop in the global topological map). In many approaches, the space consumption is similar to that of the corresponding plain representation approach as the local maps cover the complete environment.

2.3.4.5 Combining Different Organizational Forms

Aguirre & González (2002) describe a good example of a representation in which different organizational forms are combined into a complex representation structure. The proposed representation can be seen as a patchwork representation of two components. The local maps of the patchwork are overlays of a geometric representation and an occupancy grid, both describing the same local area. The global representation is formed by a two-level hierarchy of graph-based representations. The local maps are attached to the lower level of the hierarchy, which describes the environment at a fine level of granularity, while the top level provides a coarser view of the environment that is used to generate high-level plans. The representation is embedded in a hybrid deliberative-reactive architecture, and the best available representation for the task is chosen based on context, explicitly permitting a certain level of representation not to be available for a certain area. The paper, however, focuses on the application of this complex organization for path planning and execution, and does not discuss global localization in detail. In addition, the approach depends highly on a reliable detection of important environmental features like doors and corridors.

A similar representation can be found in Galindo et al. (2005). A multi-level hierarchy of graph representations is used and local grid maps are attached to the nodes of the bottom level of the hierarchy. In addition, camera images of objects are attached to nodes corresponding to the regions in which the objects have been encountered, and the spatial hierarchy is linked to a semantic hierarchy. This work nicely demonstrates the suitability of this kind of representation to interface with conceptual knowledge, allowing for high-level symbolic reasoning and planning.

Furthermore, the SSH by Kuipers (Kuipers, 2000; Kuipers & Byun, 1991; Kuipers & Levitt, 1988) already mentioned several times in this chapter describes a complex form of organization of multiple spatial representations. As it has been developed as a model of human knowledge and experience of large-scale space, it contains elements like procedural competences and a memory of sensorimotor experiences that go beyond the declarative environmental models discussed here. Nevertheless, the SSH has been realized on mobile robots several times, and organizational forms like patchworks and overlays can be identified.

The SSH distinguishes five levels of spatial knowledge and multiple representations, each using its own ontology. The *sensor level* consists of sensorimotor control laws for leading the agent to locally distinctive states of the environment. At the *causal level*, the continuous experience of moving between distinctive states based on the control laws is abstracted as view-action-view triples. View here stands for the sensory input available at a distinctive state. The *topological level* consists of a graph model that best explains the experience of the causal level. The nodes correspond to places in the environment and can be (hierarchically) grouped into regions, while sequences of edges are grouped to form paths. The nodes may be linked to local coordinate-based maps describing the neighborhood of the place in a patchwork-like manner. The *metrical level* comprises an optional overlay of a coordinate-based representation constructed from the patchwork by merging the local maps into a global frame of reference.

The SSH has been completely axiomatized and inference rules for deriving knowledge from other levels have been formulated within suitable logical frameworks, e.g., for deriving the topological level from the causal level using non-monotonic reasoning (Remolina & Kuipers, 2004).

2.4 Uncertainty Handling Approaches

After looking at the different ways proposed to represent the environment, we will now regard the robot mapping problem from the second perspective mentioned, the perspective of handling uncertainty. It has frequently been pointed out that the sources of uncertainty are numerous: limitations of the sensors and noise in the measurements, imperfect actuation, general unpredictability of dynamic environments, and the fact that models of the environment have to be coarse approximations of the real world in order to be computationally feasible (Thrun, 2000; Thrun et al., 2005). This implies that a robot can never be certain about the current state of the environment.

In the following, we make two distinctions to classify uncertainty handling approaches appearing in the literature: The first distinction concerns the nature of the update process of the spatial representation. Most real-time mapping systems (often called *online approaches*) incorporate new observations *incrementally*: The current representation is updated, leading to a new representation; the decisions made will not be revised anymore in the future. Alternatively, sensor data is not processed sequentially but a (heuristic) search for an optimal solution is performed in which previous decisions can be retracted and the input data is processed in multiple passes. As we are mainly interested in online mapping, we will focus on incremental approaches and only briefly mention the most important multi-pass techniques.

Second, we distinguish approaches based on the number of hypotheses that are being tracked simultaneously in the mapping approach. The three options here are (1) a single hypothesis only, (2) multiple discrete hypotheses, and (3) keeping track of the complete state space at any moment.

2.4.1 Incremental Approaches

In the following review of incremental approaches we pay special attention to the question of which kinds of spatial representations have been combined with which uncertainty handling approach.

2.4.1.1 Single Hypothesis Approaches

Single hypothesis approaches maintain one hypothesis about the state of the environment. Whenever a new observation becomes available, the hypothesis is updated in the most plausible way. In probabilistic approaches, in which at least the uncertainty of some individual facts is represented (e.g., a probability distribution for the current pose of the robot), this means that the most likely model is constructed, and the approach is often referred to as *incremental maximum likelihood*.

Traditionally, relational representation approaches only maintain a single hypothesis about the state of the environment. We can find this approach both in work on view graph representations (Franz et al., 1998; Schölkopf & Mallot, 1995) and in work on route graph representations (Choset & Nagatani, 2001; Kuipers, 2000; Kuipers & Byun, 1991; Kuipers, 1978; Kuipers & Byun, 1988; Kuipers & Levitt, 1988; Nagatani et al., 1998). In all of these approaches, information is represented as facts without any kind of uncertainty assessment. As relational approaches typically focus on qualitative spatial relations that can be more reliably perceived, this is less problematic than for coordinate-based representations. Nevertheless, combining observation and model in the most plausible way can be a difficult task, and often the algorithms may fail to construct the correct model under certain conditions.

Some relational approaches do maintain uncertainty information for at least some individual facts contained in the model. Examples of this are the approach of Yamauchi & Beer (1996), in which confidence values for edges in the graph-based representations are employed; the approaches by Yamauchi & Beer (1996) and Duckett & Nehmzow (1999a,b), which both explicitly represent the uncertainty in the robot position; and the approach described in Hübner & Mallot (2007), which maintains position estimates for the nodes.

Coordinate-based approaches often also only maintain a single map hypothesis. Mostly, this holds for geometric representations in which new observations are matched to and merged into the current map (Crowley, 1989; Wolter et al., 2004). Again, many approaches here at least maintain a probability distribution over the robot's pose space, and sometimes also over individual object parameters.

The advantage of only maintaining a single hypothesis is the simplicity of the approach, resulting in much lower computational cost. The downside is a general brittleness of these approaches when the input data is ambiguous. As incremental approaches do not have the ability to recover from wrong decisions made in the past, a single error in the map update step is typically fatal and results in an unusable spatial model.

To deal more adequately with ambiguous input data, one resorts to approaches that try to maintain a discrete set of spatial models. However, before we discuss these approaches, we turn to the other extreme of approaches that keep track of the complete space of hypotheses by rigorously regarding the mapping problem as a probabilistic state estimation problem.

2.4.1.2 Complete-State-Space Approaches

The common approach taken in probabilistic mapping is to maintain a probability distribution over the space of all different hypotheses and to incrementally update it based on new observations. A prerequisite for this approach is that the perception and actuation processes can be adequately modeled stochastically. The theoretical basis for updating the probability distribution is then given by the recursive Bayes filter (see Appendix A for the technical details of probabilistic mapping).

In most realistic scenarios, however, it is infeasible to represent and compute the probability distribution exactly. Therefore, approximations have to be used. During the last two decades, very powerful approximation techniques have been developed or adopted from other application areas. Most of them fall into one of two main classes: *parametric filters* and *nonparametric filters* (cf. Thrun et al., 2005).

Parametric filters—based either on the Kalman (Kalman, 1960) or the Information filter (Mutambara, 1998)—use normal distributions as approximations and have been applied in combination with landmark-based representations and geometric representations. Smith & Cheeseman (1986) were the first to use the extended Kalman filter (EKF) to formulate a solution to the landmark-based SLAM problem. This work has spurred a huge amount of research on EKF-based SLAM (Castellanos et al., 1999; Dissanayake et al., 2001; Guivant & Nebot, 2001). Besides the general inability to adequately represent multimodal probability distributions, the main drawbacks or challenges of this approach are the quadratic space and time complexities: Incorporating a new observation requires $O(n^2)$ time and the covariance matrix which needs to be maintained has a size of $O(n^2)$ (where n is the number of landmarks). In addition, the approach relies on correct data association.

In the last few years, several methods to reduce the quadratic time complexity of the EKF and Information filter have been proposed (Frese & Hirzinger, 2001; Paskin, 2003; Thrun et al., 2002). Often these approaches employ some kind of submapping strategy involving a (hierarchical) decomposition of space into subregions, e.g., the Compressed EKF (Guivant & Nebot, 2003, 2001) and Treemap (Frese, 2006b; Frese & Schröder, 2006).

Nonparametric filters approximate probability distributions by a finite set of samples. The dominant approach in this class is the so-called Rao-Blackwellized particle filter (Doucet et al., 2000), in which the probability distribution is factorized and each particle stands for a particular robot trajectory and the map associated with this trajectory. The probability of each sample is given by its importance factor. Rao-Blackwellized particle filters have first been introduced for landmark-based representations under the name FastSLAM (Montemerlo & Thrun, 2003; Montemerlo et al., 2003). A result of the factorization is that the positions of the landmarks within the map can be tracked individually, each with its own extended Kalman filter.

In Hähnel et al. (2003a) the idea of Rao-Blackwellized particle filters is first used in connection with a grid map representation. Again, the particles stand for entire trajectories. The map corresponding to a particle is computed from the trajectory using standard occupancy grid mapping. Improvements of the original algorithms and approaches to reduce the so-called particle depletion problem that arises when the number of used particles is too small to adequately represent the probability distribution are described in Grisetti et al. (2007a,b). Eliazar & Parr (2003, 2004) describe a standard particle filter approach to construct occupancy grid representations.

The approach of maintaining a probability distribution over the complete state space has been very successful in achieving a high level of robustness under uncertain conditions. Most current state-of-the-art mapping approaches fall into this class. However, as mentioned, the approach is computationally demanding and the approximation techniques still tend to have limitations or are applicable only under particular conditions. In addition, while current approaches show that a rigorous probabilistic formulation is possible for representations that are still very close to the sensor data (grid maps, landmark maps, and to a certain degree geometric representations), it is an open problem how this approach could be adapted for more abstract forms of representation and organization.

2.4.1.3 Multi-hypothesis Approaches

Maintaining a discrete set of hypotheses offers a way to reduce the brittleness of many single hypothesis approaches. Multi-hypothesis approaches can better deal with ambiguities, although they carry additional computational costs as multiple spatial models have to be maintained. At the same time, it has been seen as a way to deal with one disadvantage of parametric complete-state-space approaches: their inability to adequately represent multimodal probability distributions. Moreover, particle filter-based approaches have a kind of multi-hypothesis flavor as each particle stands for one particular hypothesis. However, the fact that they form an approximate representation of the actual probability distribution over the complete state space explains why they belong to the previous class. Two important questions in the context of multi-hypothesis approaches are about when to instantiate a new hypothesis and when to discard a hypothesis that has become implausible.

In the context of Kalman filter-based approaches, multi-hypothesis tracking techniques have been developed to deal with the ambiguities arising in the data association step, which would result in a multimodal probability distribution. Cox & Leonard (1994) employ this idea—which originally was developed in work on object tracking (e.g. Reid, 1979)—for robot mapping using a geometric representation with features for walls and corners. The approach assumes that the poses of the robot are known. This allows the use of an independent Kalman filter for each geom. Whenever there exist multiple plausible associations for the current observation, which is determined using validation gating (cf. Sect. 5.1.2), a hypothesis is split into multiple new hypotheses. Probabilities are assigned to the hypotheses and they are maintained in a hypothesis tree with the current hypotheses forming the leaves. To deal with the problem of exponential growth, the authors consider multiple pruning strategies to remove branches of the tree based on the assigned probabilities.

In Smith & Leonard (1997), this approach is extended to deal with the full SLAM problem. Hypotheses for robot pose and feature map are tracked independently (ignoring correlations between the two robot and feature states).

While in the above approaches all hypotheses are constructed whenever a new observation becomes available, and hence all leaves are at the same level of the tree, (Hähnel et al., 2003) proposes a lazy search through the hypothesis tree. In the described approach, hypotheses are only updated when they have the potential to become more likely than the currently considered hypothesis. Thus, the leaves of the tree are at different levels depending on the number of observations that have already been incorporated in the corresponding hypothesis. The lazy search is possible since the log-likelihood values of the hypotheses decrease monotonically with the depth in the tree.

In a newer implementation of Kuipers's SSH (Kuipers et al., 2004), a topological equivalent to the previously described multi-hypothesis approaches for coordinate-based representations is presented. A tree of topological map hypotheses is maintained. Every time a new view-action-view triple is experienced, all hypotheses are updated accordingly. This can lead to multiple successor hypotheses. Hypotheses that violate the axioms of the topological level of the SSH are pruned. Active exploration is used until the simplest hypothesis with respect to the logical theory has been unambiguously identified. The experiments using an environment with nine places demonstrate that the high level of abstraction of the involved representation (nodes roughly correspond to junctions and edges correspond to hallways) together with the hard constraints for pruning allow for an exhaustive search through the tree of hypotheses. While it remains open how this general approach scales to more complex environments, an investigation of how the assumption that the graph hypotheses have to be embedded into the plane affects the size of the space of hypotheses is conducted in Savelli & Kuipers (2004).

In general, the multi-hypothesis approach seems to have great potential as a compromise for achieving a sufficient degree of robustness while avoiding the computational costs of complete-state-space approaches and still being rather universally applicable, as it does not assume known data associations. The main challenge is to find good strategies to exploit all available information to reduce the search space enough to avoid the exponential growth. It is especially promising for more abstract representations for which a complete probabilistic formulation might be hard to achieve but for which the combinatorial problem is significantly reduced.

2.4.2 Multi-pass Approaches

In multi-pass approaches the input data is processed several times, allowing for revising past location estimates or recovering from wrong data associations, a property that most incremental approaches are lacking. The most well-known family of multi-pass approaches is based on a statistical approach known as *expectation maximization* (EM) (Dempster et al., 1977). Thrun et al. (1998b) use this approach to build landmark-based maps (landmarks were simulated by button presses). The idea is to search for the most likely map given the input data by using hill-climbing in the likelihood space. The EM algorithm alternates two steps, the *expectation step* and the *maximization step*. In the expectation step, a sequence of pose estimates is determined based on the current most likely map (localization with a given map), while the maximization step computes the most likely map given the sequence of poses computed in the expectation step (mapping with given pose estimates). The result of the EM algorithm can then be used, for instance, to construct a grid map from the pose estimates and sonar measurements collected in addition to the landmark information. Although naturally the approach is only guaranteed to converge to a local maximum, the experiments indicate that it performs robustly in the presence of large odometry errors. A good initial map estimate improves performance of the algorithm significantly.

Burgard et al. (1999) improve this approach and replace the manually added landmarks by local grid maps. A modified maximization step is used in which deterministic annealing is utilized in order to reduce the chance of getting stuck in a local maximum because of a bad initial map. EM-based mapping has also been applied for the plane fitting subproblem of learning compact geometric 3D models from laser range data (Liu et al., 2001).

Other multi-pass mapping techniques have been developed in the context of scan matching. Lu & Milios (1997) describe the problem of constructing a map from a sequence of scans as a global optimization problem and compute a solution based on minimizing an energy function using a spring model. Similarly, in the work of Nüchter et al. on 6D SLAM (Nüchter et al., 2004, 2005) a global relaxation approach to diffuse the accumulated error over all scans is employed in which the scans are registered multiple times with respect to their neighboring scans.

2.5 Conclusions

We have looked at the multitude of mapping approaches described in the literature, first from the perspective of spatial representation and then from the perspective of uncertainty handling. For assessing different kinds of spatial representation, we referred to the criteria proposed in Sect. 2.2 (extractability and maintainability, information adequacy, efficiency and scalability) and the three main operations of navigation, systematic exploration, and communication (Sect. 2.1).

When looking at the spatial representations employed by the approaches listed in the previous section on uncertainty handling, several major clusters become immediately apparent. For all kinds of coordinate-based representations, ways to maintain multiple or even infinitely many hypotheses, as in the complete-state-space approaches, have been developed. In contrast, relational representations in almost all cases rely on the correctness of a single hypothesis and thus tend not to work reliably when the input data is ambiguous. So far, only very few attempts have been made to develop or adapt powerful uncertainty handling approaches for relational representations.

On the other hand, as we have seen when looking at representation approaches from the spatial representation perspective, relational approaches offer advantages over coordinate-based representations that make such techniques desirable. One approach to deal with this lack of reliable construction and maintenance methods is to aim for complex forms of organization in which a coordinate-based approach is used to derive a relational representation. However, as in the case of an overlay approach combining a global coordinate-based representation with a global graph representation, this approach tends to preserve too many of the disadvantages of the chosen representations. It also seems implausible from a cognitive perspective, as humans are known to develop knowledge about the geometric layout of an environment last, and this knowledge is usually systematically distorted (Siegel & White, 1975; Tversky, 1992).

We therefore argue that developing robust methods for directly constructing and maintaining representations that are relational at the core but can be enriched with other kinds of information will lead to robot mapping systems that scale much better to larger environments and will perform much better in the context of the considered broader set of tasks. To reach this robustness, investigating ways to combine relational representations with more sophisticated uncertainty handling techniques is a necessity. The fact that relational representations typically are based on relations that can be detected more reliably should in principle make this combination even more robust than in the case of representations based on information that can be observed only very unreliably.

In this work, we follow this direction of research by investigating the combination of a particular route graph representation with multi-hypothesis tracking and robust uncertainty handling techniques. Our focus lies in developing basic methods that make this combination possible. However, we also consider and compare three different mapping systems realized based on these techniques.